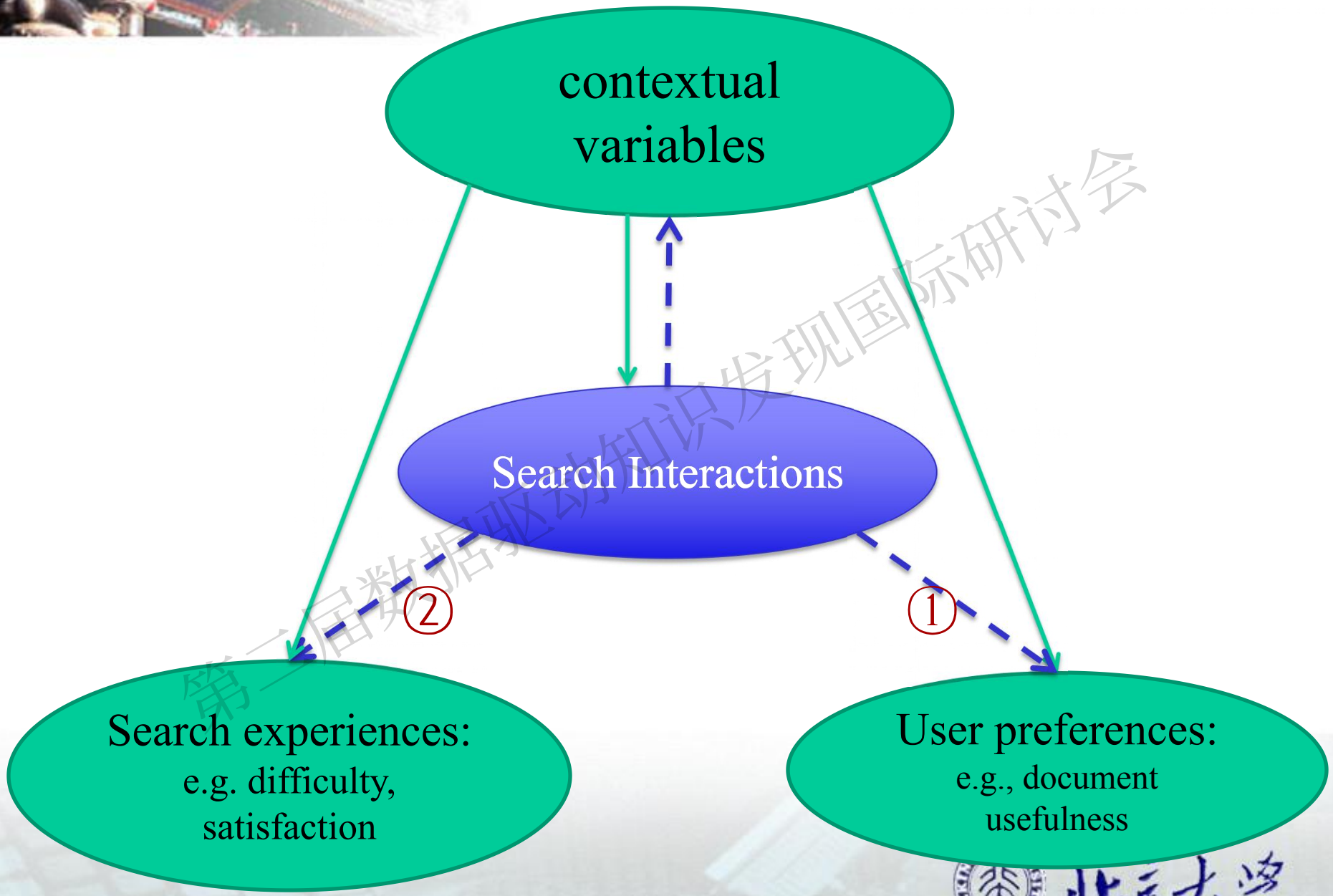
Personalization of Information Retrieval based on users' interactions and search contexts

Chang Liu

Peking University

北京大学

Interactive IR



Predicting users' preferences in different task types

Addressing the questions: User experiment

- 32 participants
- Four different search task types
- Pre-task questionnaire
  - familiarity with the search task and topic, and the estimated search difficulty
- During Search
  - Searchers' interactions with computers were logged
  - Users' preferences of pages were recorded
- Post-task questionnaire
  - experienced difficulty of the task and satisfaction with performance of the task

Facets of task which were varied in the experiment

Facets	Values	Operational Definitions/Rules
Product	Factual	Locating facts, data, or other similar items
	Intellectual	Produces new ideas or findings on the basis of locating facts
Goal (quality)	Specific goal	Goal is explicit and measurable
	Amorphous goal	Goal cannot be measureable
Naming	Named	Locating factual information about named fact
	Unnamed	Locating factual information about unnamed fact
Level	Document	Judgment is made on the document as a whole
	Segment	Judgment is made on part(s) of a document

(Li & Belkin, 2008).

Four search tasks

Task	Product	Level	Goal (Quality)	Objective complexity
Background Information Collection **(BIC)**	Mixed	Document	Specific	High
Copy Editing **(CPE)**	Factual	Segment	Specific	Low
Interview Preparation **(INT)**	Mixed	Document	Mixed	Low
Advance Obituary **(OBI)**	Factual	Document	Amorphous	High

Search tasks

- **Copy Editing (** Factual, Low complex, Specific goal
- **Your assignment:** You are a copy editor at a newspaper and you have only 20 minutes to check the accuracy of the three underlined statements in the excerpt of a piece of news story below.
- New South Korean President Lee Myung-bak takes office
- *Lee Myung-bak is the **10th** man to serve as South Korea's president. He pledged to make the economy his top priority during the campaign. Lee promised to achieve 7% annual economic growth, double the country's per capita income to **US$4,000** over a decade. Lee, **66**, also called for a stronger alliance with top ally Washington and implored North Korea to forgo its nuclear ambitions and open up to the outside world, promising a better future for the impoverished nation*.
- **Your Task:** Please find and save an authoritative page that either confirms or disconfirms each statement.

Search tasks

- **Advance O**[illegible] Factual, High complex, Amorphous goal
- **Your assignment**: Many newspapers commonly write obituaries of important people years in advance, before they die, and in this assignment, you are asked to write an advance obituary for a famous person.
- **Your task**: Please collect and save all the information you will need to write an advance obituary of the artist Trevor Malcolm Weeks.

Behavioral measures as sources of implicit relevance feedback

Behavioral measures on clicked documents

How does the user interact on the clicked documents?

- *dwell time*
- *number.of. mouseclick*
- *number.of. keystrokes*
- *visit_id*

Behavioral measures during query intervals

What does the user do between issuing one query and the next?

- *time_to_first_click*
- *content_mean*
- *content_sum*
- *content_count*
- *serp_mean*
- *serp_sum*
- *serp_count*
- *prop_content*
- *interval_time*
- *diff_content*

Model generation methods

- Predictive models
  - General model (non-task specific)
  - Specific model (task-specific)
- Model generation methods
  - Recursive partitioning (RPART)-decision tree model
  - Logistic regression model- LR equation model
- Four models
  - General_tree & General_LR model
  - Specific_tree & Specific_LR model

Results: general vs. specific model

- The task-specific predictive model had better prediction accuracy than the general predictive model.

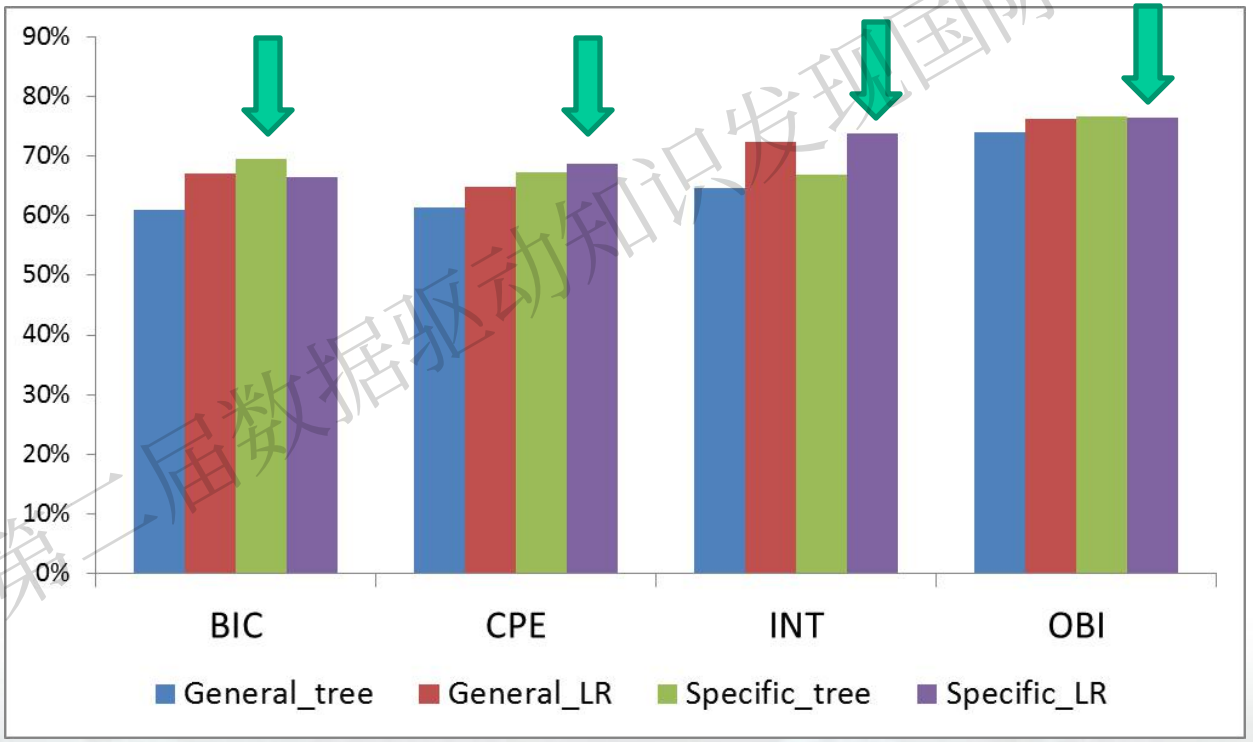
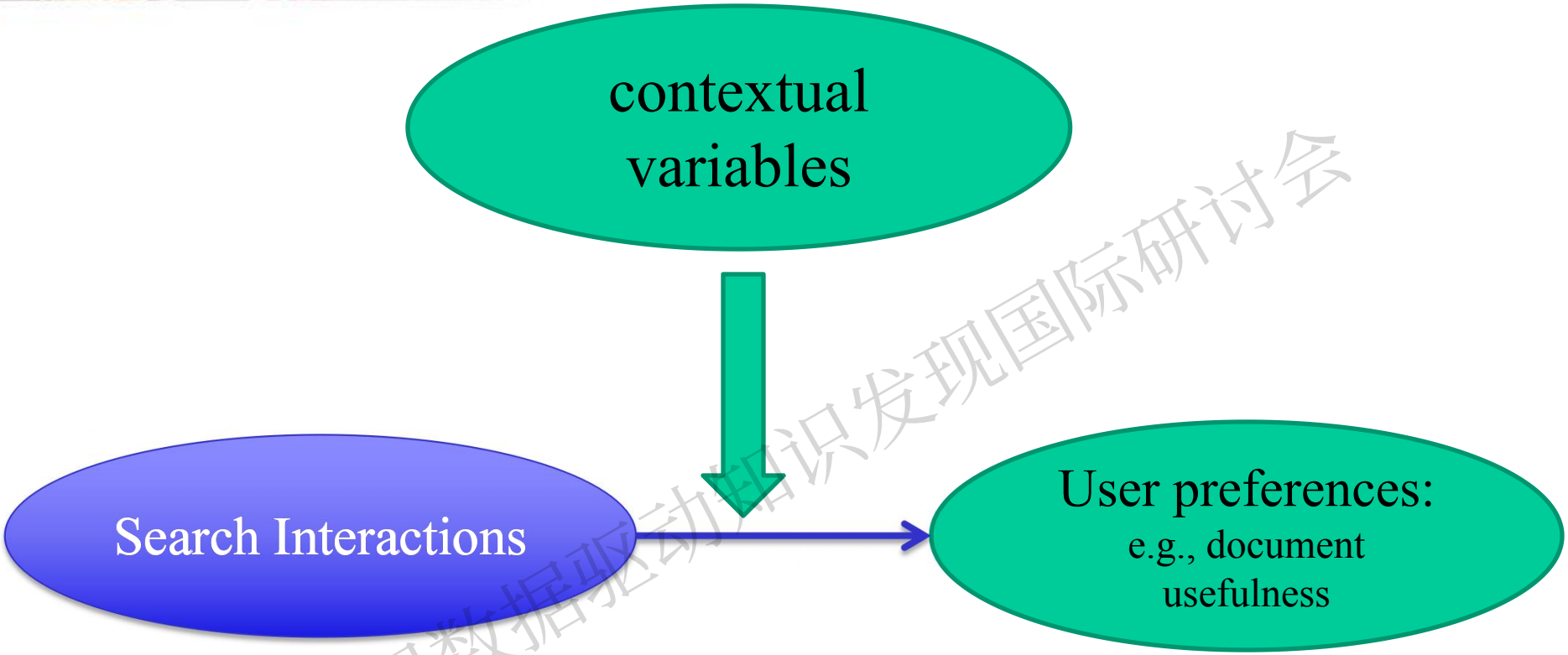


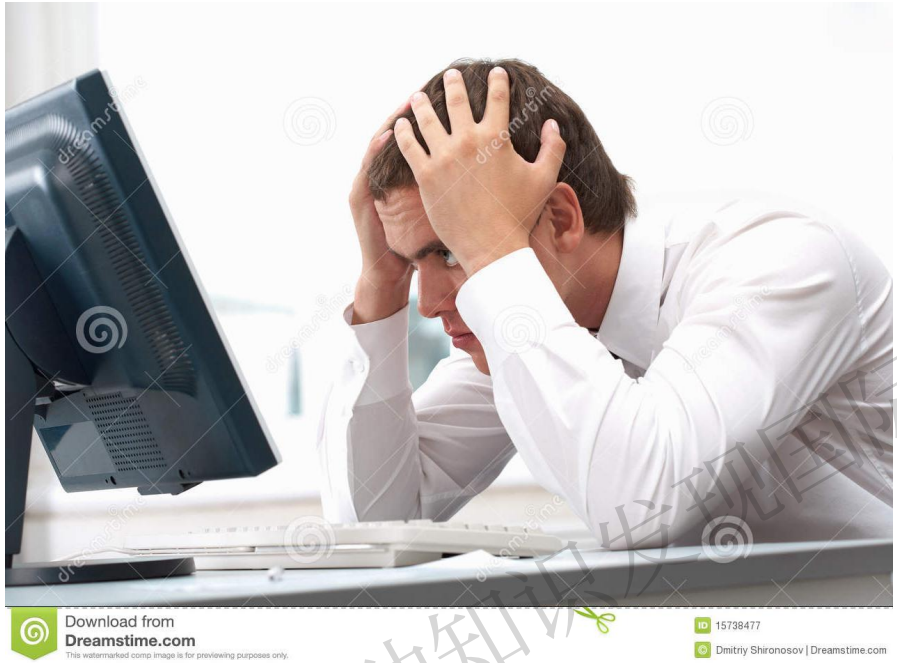
Figure. The comparison of general and specific predictive models on the overall accuracy

Conclusions

- **Task type is important!**
- If the task type information is known by the system, then it is possible to generate better predictive models.
- Users may need system assistance in certain types of tasks.

Interactive IR





Predicting Search Task Difficulty at Different Search Stages

Research Goal

- Goal:
  - Predict when searchers are having difficulty through observation of users' interactions in real-time during search sessions.
  - Generate predictive models of task difficulty at three different search stages
    - at the beginning of search
    - in the middle of search
    - at the end of search

Search task difficulty and user behaviors

- A subjective measure
- Reasons for task difficulty
  - task features
  - user aspect
  - user-task interaction
- Behavioral differences
  - difficult tasks had longer task completion time, more diverse queries, advanced operators, more queries, longer dwell time on content pages and SERPs, etc.

Task difficulty prediction

- Task difficulty, frustration, search success, search engine switch, etc.
- Whole-task-session
  - can only be acquired at the end of search
  - not helpful for real-time prediction
- Earlier stage in the session
  - First-round and accumulated level
  - After 2nd query, 3rd query…

We added the middle point of search to make the predictions of task difficulty

Search stage in information retrieval

- Searchers' feelings, thoughts, and actions vary along the different stages (Kulthau, ISP model, 1991)
- Users' criteria of assessing the relevance of Web pages for information seeking tasks through the duration of a task displayed a certain degree of variation (Tombros, Ruthven, & Jose, 2004)
- Task stage helped interpret document usefulness from the first dwell time on documents (Liu & Belkin, 2011)

Research questions

(1)What are the differences in behavioral variables between easy and difficult tasks at different search stages?

(2)What behavioral variables are important predictors of task difficulty at each of the search stages?

(3)Which search stage could achieve the best prediction performance?

(4)Are there any differences in prediction performance at different search stages by task type?

Variables in this study

- Task difficulty levels evaluated by searchers
- Behavioral measures as predictors
- Search stages

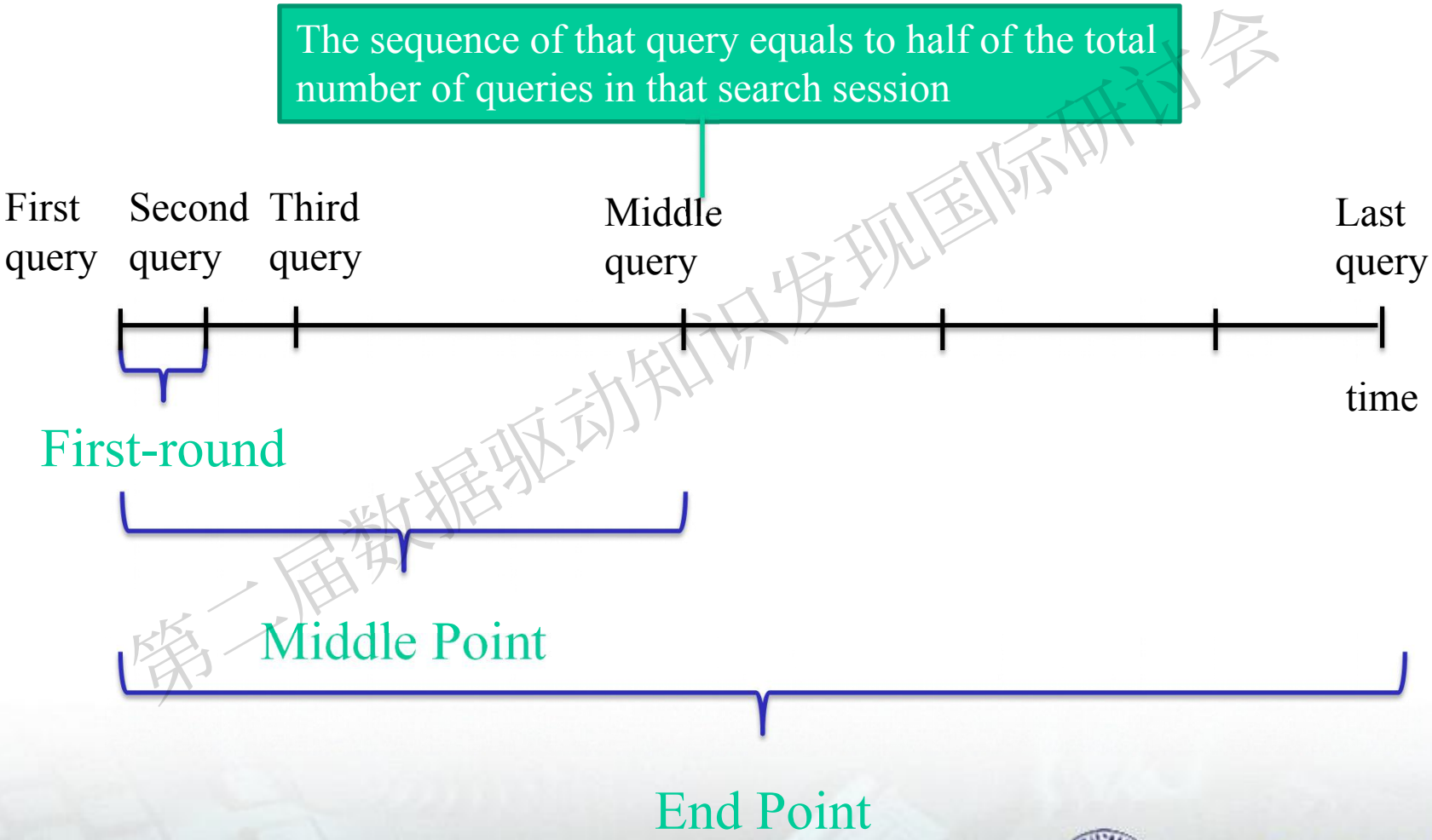
Task Difficulty

- In post-search questionnaire:
  - how difficult they thought the tasks were after completing the search tasks
  - based on a 7-point scale
- The scores were collapsed for data analysis:
  - *Difficult* for user ratings of 5-7;
  - *Neutral* for a rating of 4;
  - *Easy* for ratings of 1-3.

Behavioral measures

- *Average query interval time*
- *Mean dwell time of all documents*
- *Mean dwell time of all SERPs*
- *Average total dwell time of unique documents*
- *Average total dwell time of unique SERPs*
- *Number of viewed documents per query*
- *Number of unique viewed documents per query*
- *Number of viewed SERPs per query*
- *Number of unique viewed SERPs per query*

Three search stages



Features by tasks

Task	BIC	CPE	INT	OBI
Objective Complexity level	High	Low	Low	High
Average rating of difficulty	4.53	2.31	3.31	5.25
Average task completion time (min)	18.08	9.45	11.78	12.54
Average # of unique pages visited	25.44	10.66	21.31	20.53
Average number of queries	20.79	7.86	12.71	17.48

Behavior comparisons- First Round

Behavioral measures	Task Difficulty level	
	Easy (Mean)	Difficult (Mean)
Average query interval time	**46.39**	**32.13**
Mean dwell time of all documents	**17.03**	**8.67**
Average total dwell time of unique documents	**15.36**	**9.79**
Mean dwell time of all SERPs	9.31	11.16
Average total dwell time of unique SERPs	15.67	16.26
Number of viewed documents per query	**1.51**	**0.9**
Number of unique viewed documents per query	**1.28**	**0.75**

Behavior comparisons- Middle Point

Behavioral measures	Task Difficulty level	
	Easy (Mean)	Difficult (Mean)
Average query interval time	**56.3**	**47.09**
Mean dwell time of all documents	17.02	13.61
Average total dwell time of unique documents	20.5	17.49
Mean dwell time of all SERPs	9.97	10.07
Average total dwell time of unique SERPs	17	17.61
Number of viewed documents per query	1.93	1.8
Number of unique viewed documents per query	1.47	1.39

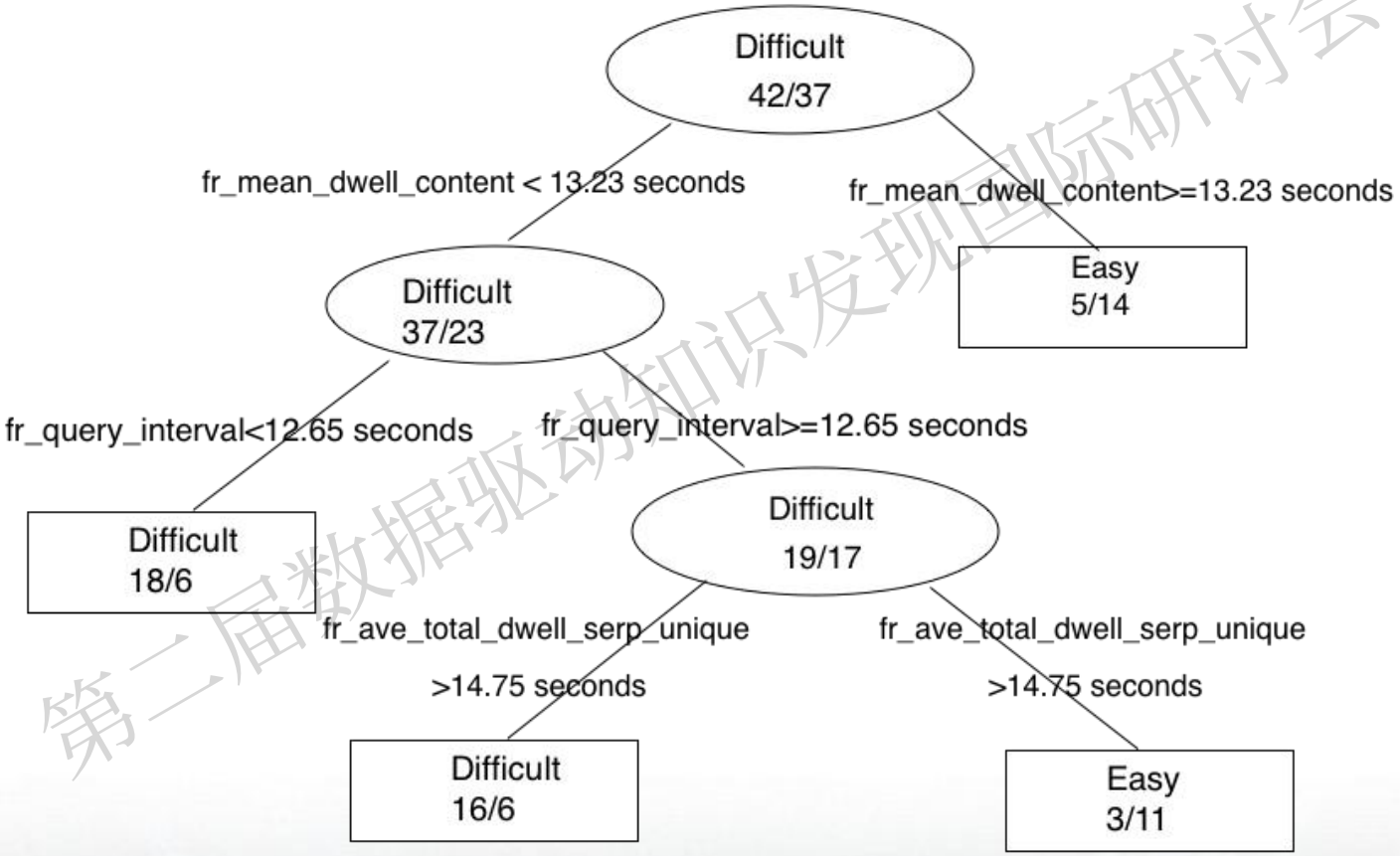
Behavior comparisons- End Point

Behavioral measures	Task Difficulty level	
	Easy (Mean)	Difficult (Mean)
Average query interval time	**63.81**	**48.09**
Mean dwell time of all documents	**17.68**	**12.36**
Average total dwell time of unique documents	**20.92**	**16.58**
Mean dwell time of all SERPs	9.67	9.8
Average total dwell time of unique SERPs	17.56	17.05
Number of viewed documents per query	**2.64**	**2.06**
Number of unique viewed documents per query	**1.84**	**1.53**

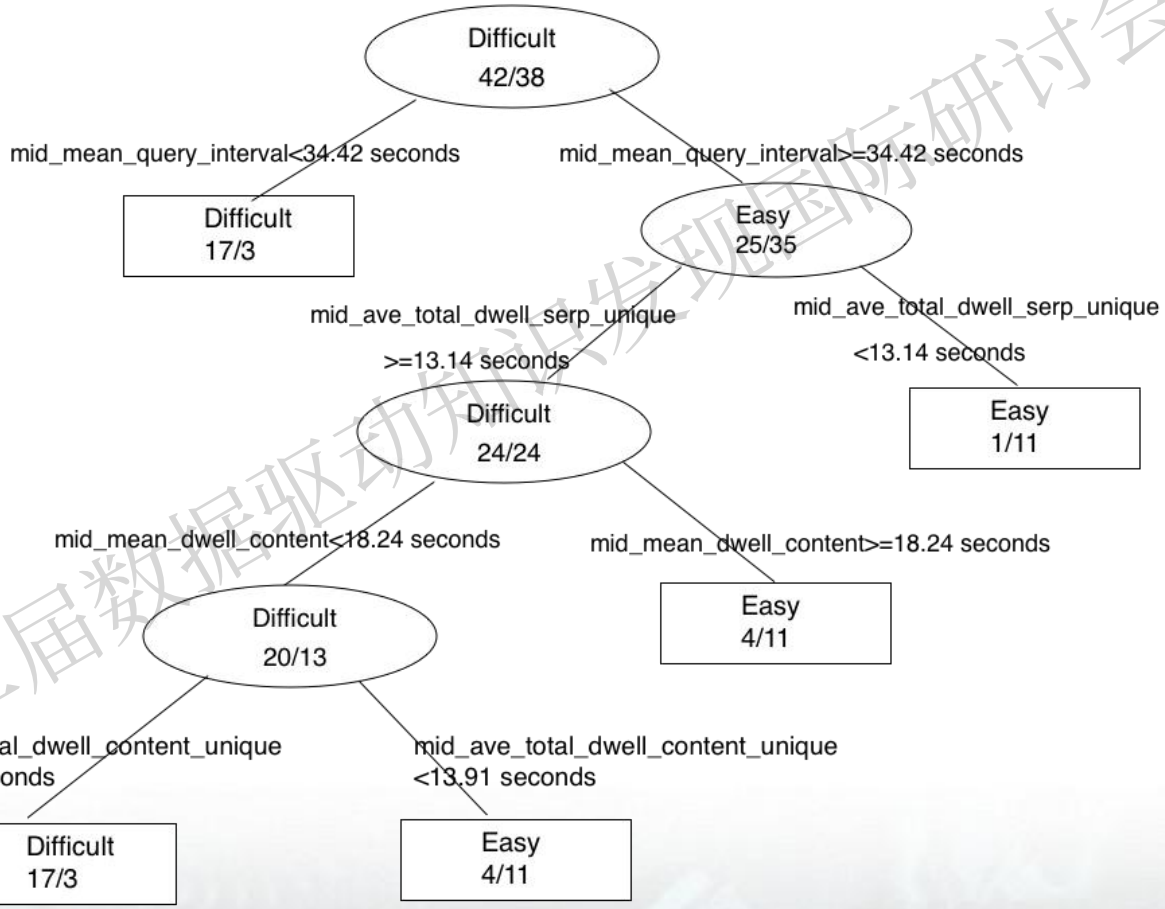
Model generation method

- Recursive partitioning (RPART)
  - Decision tree model
  - Advantage: robust to non-normal data and variable correlations
- Model generation on training sets
  - 5-fold cross-validation
  - one prediction model on each training set
  - the performance of each model is the average across all 5-test sets

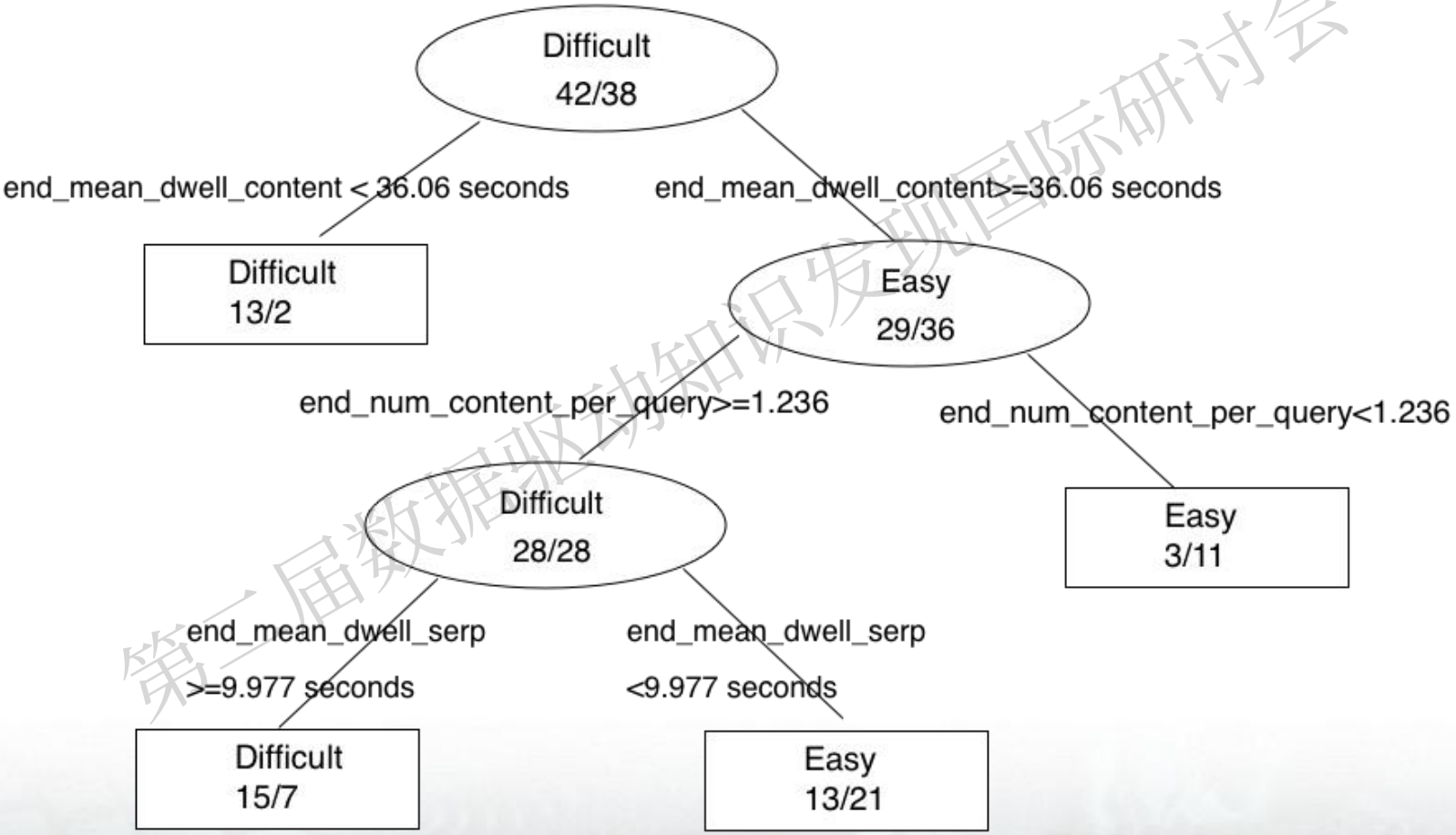
First-round Model



Middle-point Model



End-point Model



Behavioral predictors

Stage	Behavioral variables as predictors
FR model	• *Average query interval time* • *Average total dwell time of unique documents* • *Mean dwell time of unique SERPs*
MI model	• *Average query interval time* • *Average total dwell time of unique documents* • *Mean dwell time of all documents* • *Average total dwell time of unique SERPs* • *Number of viewed documents per query* • *Number of viewed SERPs per query*
ED model	• *Average query interval time* • *Average total dwell time of unique documents* • *Mean dwell time of all SERPs* • *Number of viewed documents per query* • *Number of viewed unique documents per query*

Prediction performance

Model	Accuracy	Precision	F (β=0.5)
Baseline	52.5%	52.5%	58.0%
FR model	**60.7% (+15.6%)**	61.0% (16.2%)	**62.6% (+7.9%)**
MI model	54.5% (+3.8%)	59.9% (+14.1%)	55.8% (-3.8%)
ED model	54.6% (+4.0%)	**61.6% (+17.3%)**	55.9% (-3.6%)

FR model had the best prediction accuracy and F score;
ED model had the best precision of difficult tasks;
MI model was not as good as FR and ED models.

Prediction accuracy by task type

Task	FR model	MI model	ED model
BIC	**64%**	46%	50%
CPE	68%	59%	**82%**
INT	38%	**63%**	42%
OBI	**72%**	52%	48%

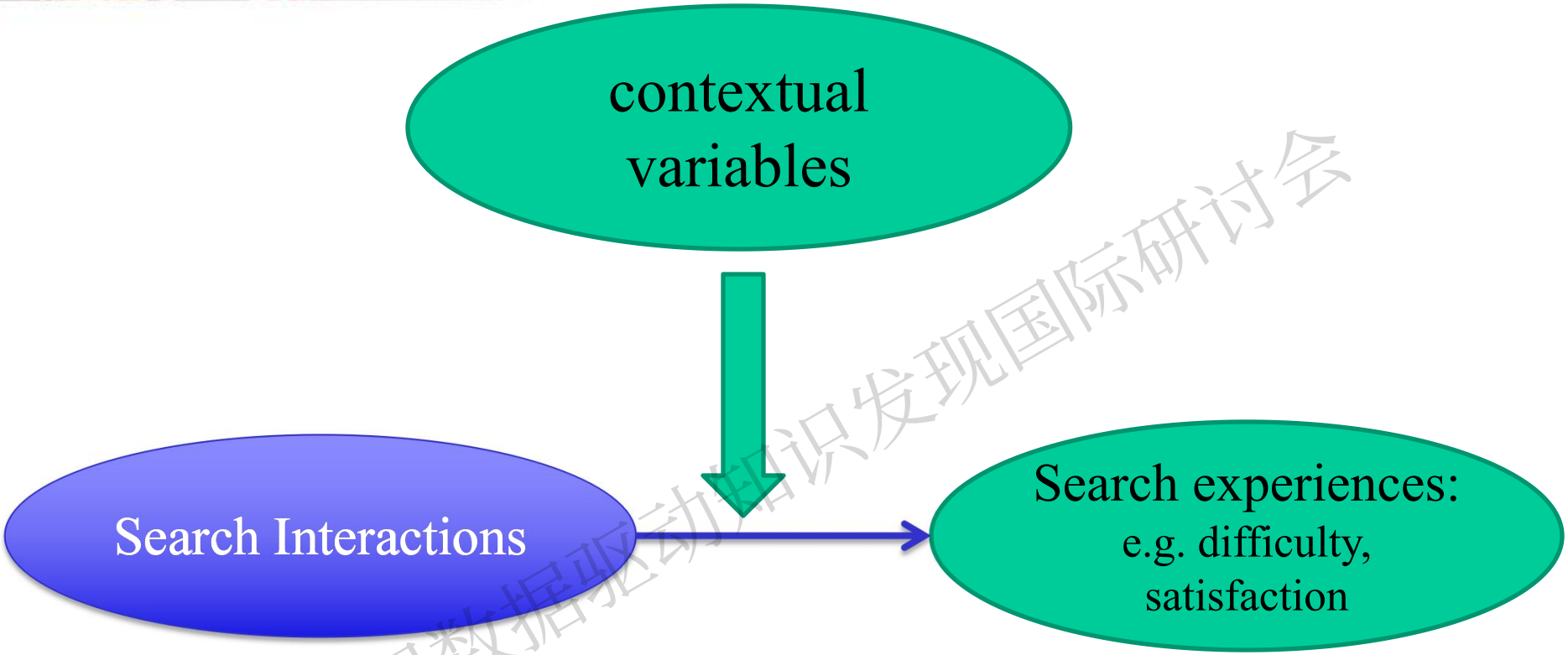
Model comparison by task:

- FR model: BIC & OBI (more complex and more difficult)
- MI model: INT (two parts of searching, intellectual search)
- ED model: CPE (factual task, relatively simple and easy)

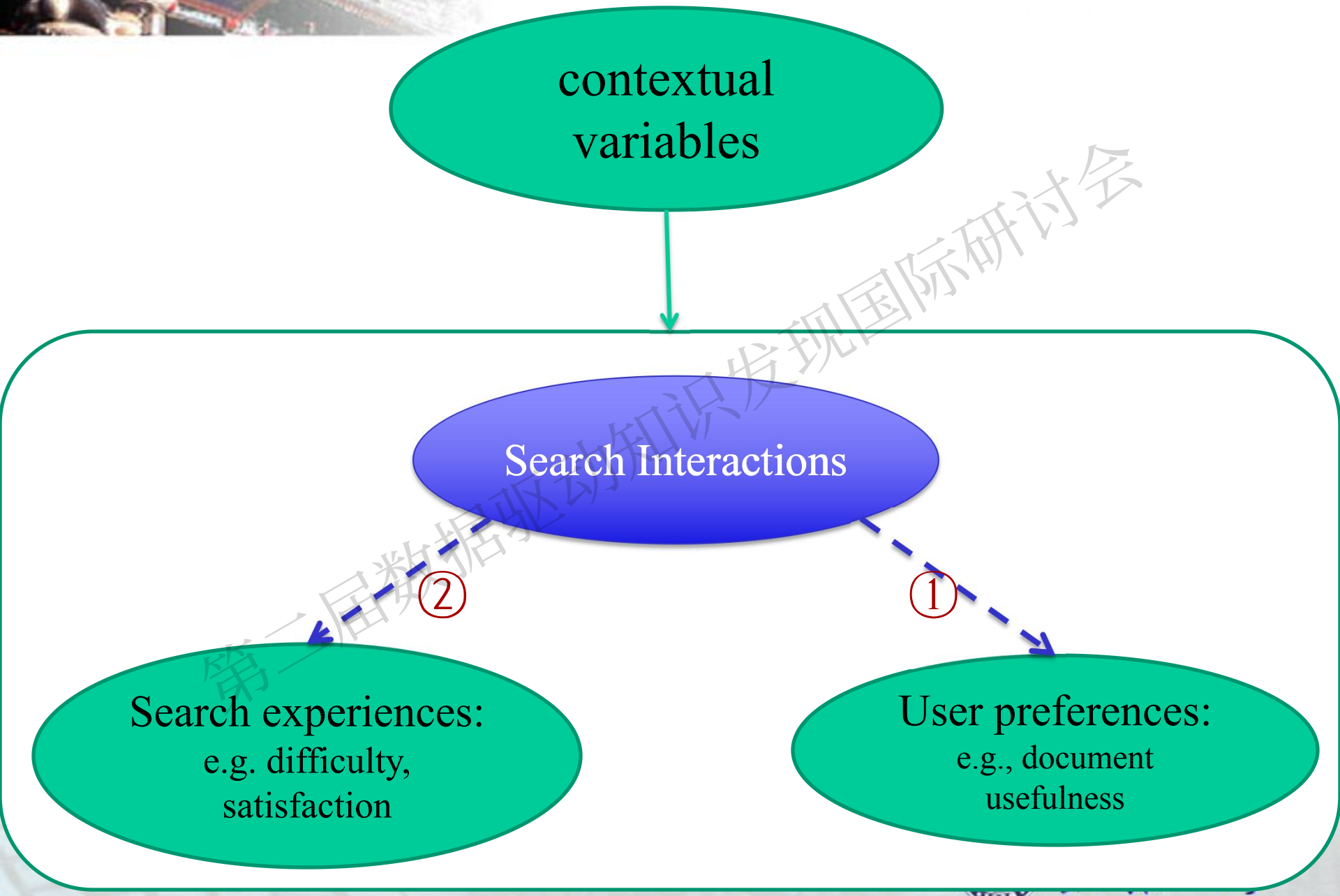
Summary

- Using a limited number of variables in the predictive model can obtain relatively good prediction performance
- FR model had the best prediction performance
- First-round of search is an important point to make predictions of task difficulty

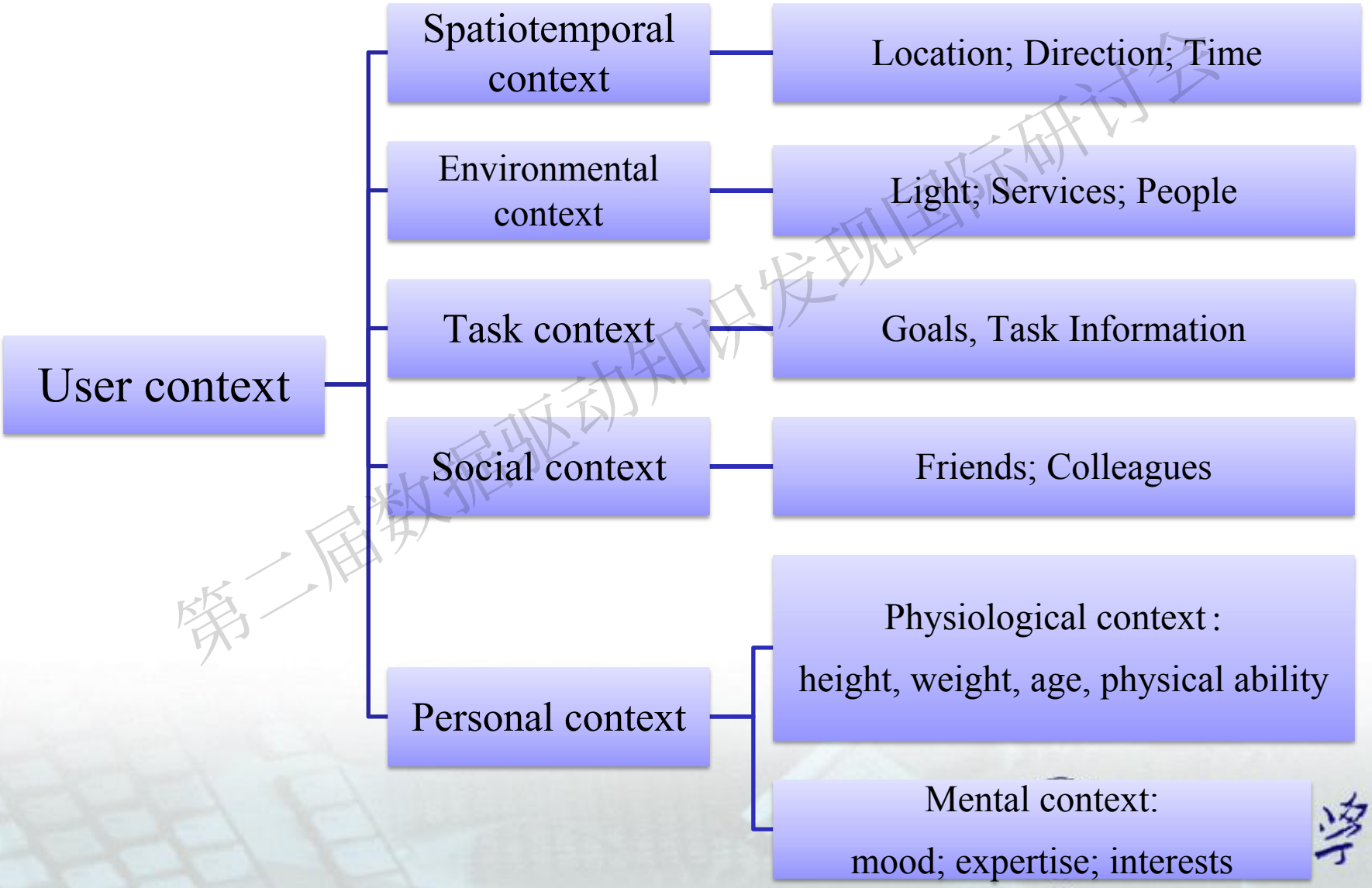
Interactive IR



Interactive IR



Search Context



Thank you!

- Questions or Comments?

Contact information:
Chang Liu, imliuc@pku.edu.cn

北京大学